Analyze building the residential buildings worth it by cart algorithm

Mingqian Zhang[1*]
[1]*Department of Mathematics and Computer Science, Denison University, Ohio, 43023, United States*
Zezhou Ma
[3]*Temple City High School, California, 91780, United States*
Feng Chang[5]
[5]*College of Big Data and Software Engineering, Zhejiang Wanli University, Ningbo, 315100, China*
Xinzhuo Yan[2]
[2]*School of International Education, Beijing University of Chemical Technology, Beijing, 100029, China*
Richard Zhang[4]
[4]*Shanghaitech University, Shanghai, 200000, China*
Chenyu Fang[6]
[6]*Earl Haig Secondary School, Toronto, M2N 3R7, Canada*
[*]*Corresponding author Email: 787368316@qq.com*

Abstract: Forecasting house prices is important for predicting economic trends in each country. In addition, building a house is a huge expense for both the government and real estate agents, so agencies need to be more careful in deciding whether to build or not. As a result, construction costs and home sales prices will be a crucial determinant. In this article, different attributes of building indicators affect the selling price and construction cost of housing were analyzed. First, a visual analysis of some variables in the dataset and discussed their relationship with sale price and construction cost was made. feature engineering came second. Based on some economic knowledge and the correlation between variables and outputs, variables were screened out to better predict sale price. Finally, a decision tree was used to test and get a model that can predict sale price most effectively.

Keywords: residential building, Iran, construction cost, sale price, correlation calculation, CART algorithm

1. Introduction

It is widely known that the COVID-19 pandemic is drastically affecting the entire world. Its impact is also reflected in the real estate market, meaning that predicting housing prices is important for forecasting the trend of the economy in every nation. In addition, building houses will be a great expense for both the government and the real estate agency. Due to the unstable economic condition recently, agencies need to make decisions on whether to build or not more carefully. Therefore, the construction costs and the sale prices of the houses would be a crucial determinant. Previous research about how different attributes affect the sale price and construction cost of the house provide this project with many ideas. Chen Bingyu published an article (2021-03) which admitted that the interest rate and loan will have a great impact on the prices of the house [1]. Zhang Ming and Liu Yao's article (2021-06) proved that the lot area, distance from CBD, cumulative liquidity, population and its structure will have a combined impact on the sale price of the house [2]. What's more, Pan Ting (2021-02) also analysed the influencing factors of house price from four aspects: demand, supply, policy and customer [3]. Waiwai published an article (2017) about the prediction of the sale price of the buildings on Zhihu[4]. He used xgboost and linear regression model to deal with 9 physical features (overall material and finish quality, total kitchens above grade, above grade (ground) living area square feet, enclosed porch area in square feet, size of garage in square feet, total square feet of basement area, original construction date, overall condition rating, the building class) of the building to make prediction and finally get a model with 90% accuracy. He firstly cleaned the dataset by examining the missing data and delete the related data. And then he calculated the correlation between the features and the sale price and decided on the nine features mentioned above to build the prediction model. Zhengyang(2018) built a random forest model to predict the sale prices of the houses. He also cleaned the data and did feature engineering to find out which feature is valuable to be used in the model. Zhengyang built the model using a grid search method and finally got a model whose accuracy is about 0.86[5]. The features used in the model are physical locations within Ames city limits, exterior material quality, height of the basement, interior finish of the garage, garage location and type of foundation, which are also physical features.

This paper presents a novel and comprehensive model for estimating the construction cost and sale price of residential buildings in any nation at the design phase or at the beginning of the construction. This research is aimed at analyzing both the physical and financial features of the building and the economic features during the construction and build a precise model using a decision tree.

2. Data Structure

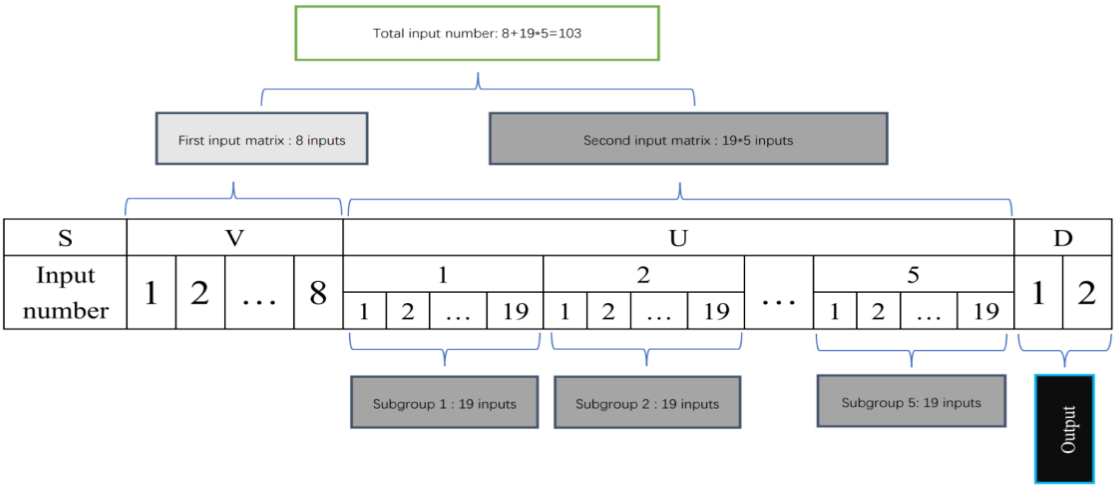


Figure 1. General data structure for the dataset

The general data structure is shown in Figure 1 above. The dataset has 2 sets of inputs and 2 outputs. The outputs are the construction costs and the sale prices of the residential building, on which the model makes predictions. The first set of inputs contains 8 physical and financial variables, and some of them have a great impact on the result of the prediction. The second set of inputs consists of 5 subgroups representing 5 periods prior to the start of the construction. The period can be a quarter, a month or a week depending on the resolution of the available data. Each subgroup has 19 economic variables affecting the price of the unit permanently.

Table 1 and 2 present the content of the first set of the input and the second set of the input respectively. Some of these parameters are for a preselected year (i.e.,1383). These variables are essential for the prediction because prices/indices vary from year to year due to inflation, deflation, or other economic conditions. According to previous research, all these variables will influence the outputs to different degrees.

The dataset had a total of 103 input variables, which was too large for modelling work. Also, such many inputs will make the training of the model computationally very intensive, which is known as the dimensionality curse. So, this project followed the example of some previous work and calculated the correlation between the inputs and outputs. The feature engineering work helped to choose several important inputs to build the prediction model.

Table1: Physical and Financial Properties for Residential Buildings (First Input Set V) [6]

Input number	Descriptions	Unit
V-1	Project locality defined in terms of zip codes	N/A
V-2	Total floor area of the building	
V-3	Lot area	
V-4	Total preliminary estimated construction cost based on the prices at the beginning of the project	10000000 IRR
V-5	Preliminary estimated construction cost based on the prices at the beginning of the project	10000 IRR
V-6	Equivalent preliminary estimated construction cost based on the prices at the beginning of the project in a selected base year (1383)	10000 IRR
V-7	Duration of construction	quarter
V-8	Price of the unit at the beginning of the project per	10000 IRR

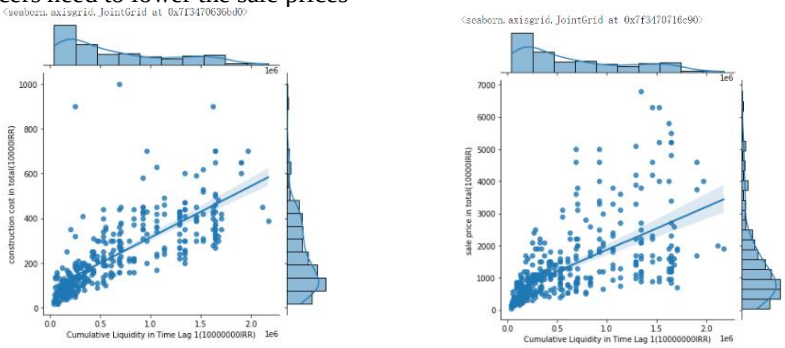
Table2: Sample Economic Variables/Indices for Residential Buildings (Second Input Set U)[6]

Input number	Descriptions	Unit
V-15	Cumulative liquidity	10000000 IRR
V-20	The interest rate for loan in a time resolution	%
V-21	The average construction cost of buildings by private sector at the time of completion of construction	10000 IRR/
V-22	The average of construction cost of buildings by private sector at the beginning of the construction	10000 IRR/
V-23	Official exchange rate with respect to dollars	IRR
V-24	Nonofficial (street market) exchange rate with respect to dollars	IRR

3. A Preliminary Exploration

3.1 Visualization

First, this work observes the relationship between construction cost and sale price and cumulative liquidity. Cumulative liquidity refers to the representation of how quickly different types of assets can be changed to cash. In other words, liquidity determines how quickly assets can be sold and whether they are priced above or below market value. Property that is easily sold and purchased at market value is liquid. Conversely, assets that are hard to sell and trade at a discount are considered illiquid. Why do buyers and investors care about liquidity? This is because liquidity determines how successfully participants exit the project. If houses are illiquid, then it is unlikely to attract buyers as they are hard to sell. Thus, producers need to lower the sale prices in order to be competitive in the market. At the same time, producers want to limit construction costs to make up for lost profits due to low liquidity. This work proposes an assumption that both the construction cost and sale price per square meter will fall as cumulative liquidity is low. However, these graphs do not exactly show what we predicted as it is illustrated in Figure 2. Looking at the plot on the left, this work concludes that the trend follows the assumption that the cumulative liquidity and the construction cost have a positive relationship. But for the plot on the right, it is more like a normal distribution in which liquidity and sale price per square meter positively correlate before the cumulative liquidity is around 1 10000000IRRm and then the relationship reverses.



(a) **(b)**

Figure 2 (a) Construction cost and Cumulative liquidity **(b)** Sale price and Cumulative liquidity

Next, this work assumes that profit that is calculated by subtracting construction cost from sale price and duration of construction are positively correlated because producers might ask for higher sale prices when the duration of construction is long enough. However, the scatter plot does not support the assumption. The plot is shown in Figure 3. When the duration of the construction is too short or too long like 1 quarter or above 10 quarters, the profit per square meter will be low. In addition, when the duration of the construction is between 5 quarters and 10 quarters, the profit per square meter is higher. However, it is not surprising to observe that a shorter period of construction will lead to low profit. An extreme short duration of construction means that this residential project was a simple and effortless one which is likely to earn less profits. On the demand side, the consumers are unwilling to pay high prices for buildings having short construction time because it is reasonable for them to doubt the quality and utility of the house. However, a residential building with an extremely long construction time is not profitable for producers. It is possible that there are

construction delays and blockers. Construction delays occur when things don't go according to plan, which is usually unavoidable. Some of the delays are caused by errors from contractors and subcontractors. These delays are within the control of the project party. Delays caused by weather and other adverse conditions are beyond the control of the project party. For example, construction work cannot continue because of snow and freezing temperatures. However, blocker is a task that must be completed before any other task can begin. For example, you want to brick a building, but you wouldn't be able to do so if the foundation wasn't built or watered [7]. These factors that interfere with construction are making the residential buildings less profitable. In all cases, construction costs and sale prices are low, even with long construction times. It can be seen by graphing construction cost and sale price with duration of time. Two scatter plots are shown in Figure 4. They follow the same pattern with the previous scatter plot that is shown in Figure 3.

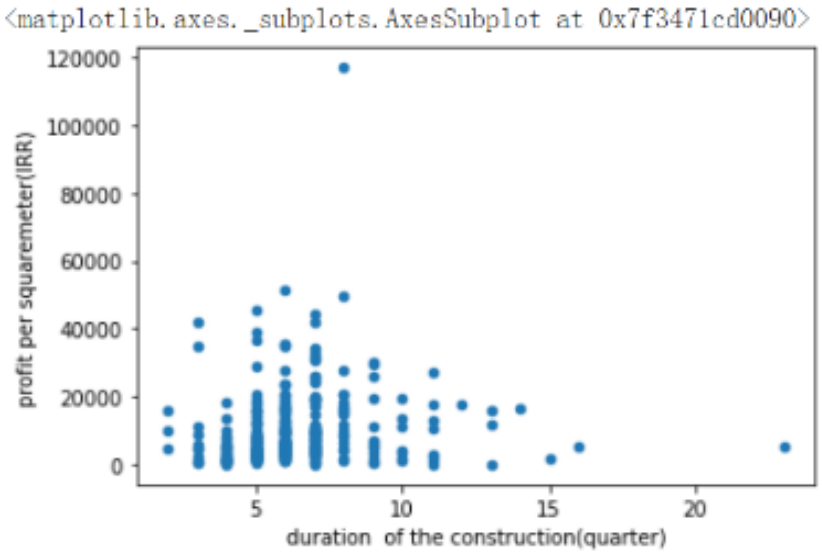
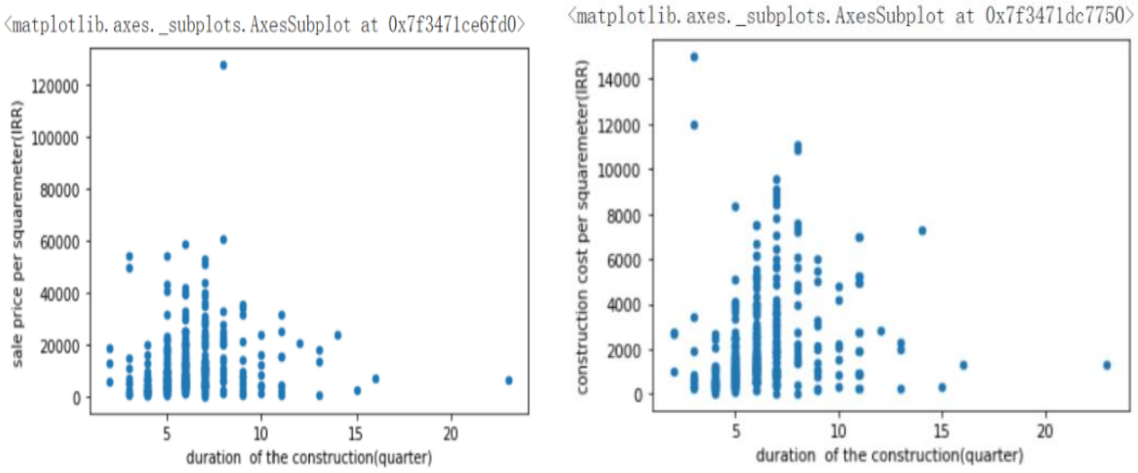


Figure 3. Profit per m2(IRR) and Duration of the construction (quarter)



(a) **(b)**

Figure 4 (a) Construction cost per m2 and Duration of the construction (**b)** Sale price per m2(IRR) and Duration of the construction

After looking at data that is interesting and might be a good predictor of our outputs, this work also ran the correlations of each variable with both outputs and found that two independent variables had extremely high correlations with them, respectively. The correlation table is shown in Figure 5.

Figure 5. Correlation of features (physical and financial variables)

First, the correlation of construction cost and preliminary estimated construction cost based on the prices at the beginning of the project is 0.96. When available information is limited, a preliminary estimate is often made at the beginning of the project. It allows trading or home-service companies to calculate project budgets and what to charge customers. The graph illustrated in Figure 6 shows a positive relationship between them, and it is obviously not surprising to us because producers and contractors can use information from similar projects done in the past to estimate the actual construction cost based on the prices at the beginning of the project. Most points are close to the regression line which means that they make good predictions of the actual construction cost based on the information at the beginning of the project. However, there are few outliers. As the article mentioned earlier, the information available to estimate costs is limited at the beginning of a project, and preliminary estimates are usually rough and approximate, giving only a rough idea of how much a project will cost. Therefore, there are likely big variations during the construction that make the prediction inaccurate. For example, the point pointed at by the arrow on the graph shows that the actual construction is underestimated by the preliminary estimate. There might be unforeseen expenses during the construction phase due to lack of experience in project leadership, poor communication among staff, loss due to damage or failure, improper allocation of resources and so on [8]. So, it is important for contractors to set a budget that is slightly higher than the estimated cost.

If a contractor really wants to decide whether to build a residential building based on the estimate made at the very beginning of a project, the preliminary estimate is not the best one. There are detailed estimates and quantity estimates which are all based on more detailed quantitative information. A company can make a detailed estimate based on a preliminary estimate. Detailed estimates provide a great deal of information about volumes, costs and rates. In addition, detailed estimates include information on the rates used to calculate the costs, specifications, drawings of the area covered by the project and detailed estimates, which are often used in the contractor's estimates. According to Carnegie Mellon University, the planning tool helps contractors know how much cash flow they need and whether they need financing in addition to figuring out what a project will approximately cost and what to charge the client. The quantity estimate should show all the materials needed to complete a project and their unit prices. So, this is a key estimate in construction. Cost figures are calculated by multiplying the dimensions on the project drawings by the rate of work on a particular project. The benefit of quantity estimation is that it can help various stakeholders understand the cost impact of their proposed changes at all stages of the construction process, thereby limiting budget overruns due to project modifications. Bid estimates are more complicated than detailed estimates and quantity estimates [9-11].

The sale price per square meter and price of the unit at the

beginning of the project per square meter have correlation value 0.98. The regression line in Figure 7 gives us a sense that they are positively correlated. But if you look closely at these points, they are centralized before 2000 10000IRR, and then the points start to get very scattered around the regression line. Especially, the price of the unit at the beginning of the project per square meter is generally greater than the sale price per square meter after 2000 10000IRR. It can be explained by the size of the residential buildings. If you have a large floor area and need a lot of building materials, manufacturers will usually buy materials in bulk or at wholesale. This way, you can decrease the price of the unit during the construction per square meter which makes the sale price per square meter lower. The article did some research on the background of Iran during the 1970s to 1980s and found that wars are likely to make residential buildings unattractive since demand for a home was low due to a turbulent economic and political backdrop [12]. The article also looked at the population of the target city and the number of building permits issued. The number of building permits grew much faster than the population of the city which means that the supply of the residential buildings is much greater than the demand. It creates a surplus and producers are likely to lower the sale price or sale price per square meter to attract consumers.

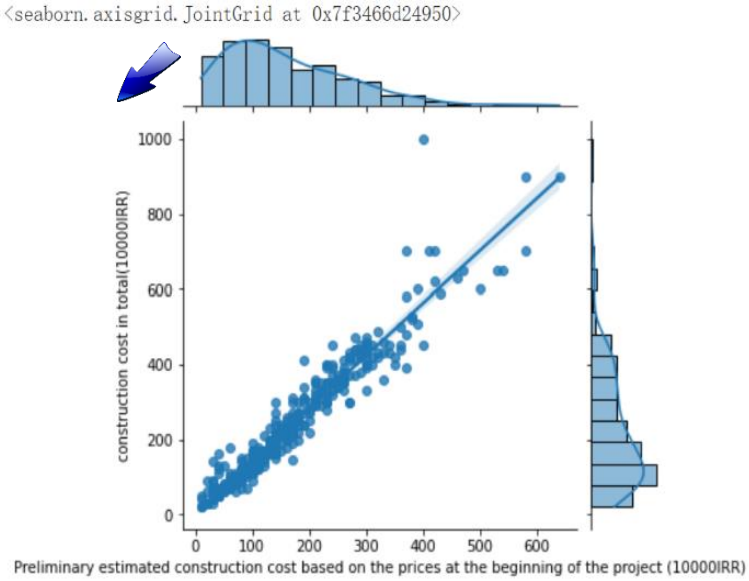


Figure 6. Construction cost and Preliminary estimated construction cost based on the prices at the beginning of the project

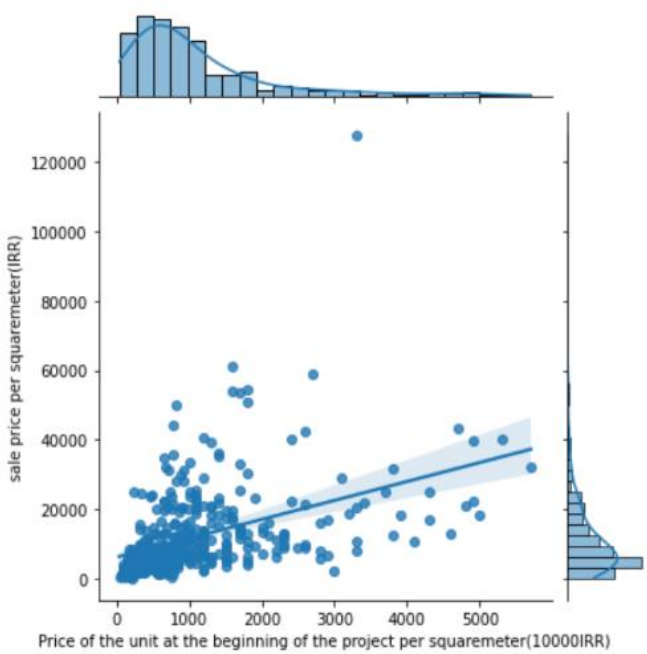


Figure 7. Sale price per m2 and Price of the unit at the beginning of the project per m2

Other points in Figure 7 demonstrate that the sale price per square meter is costlier than the price of the unit at the beginning of the project per square meter. This can happen when residential floor space is too small to save unit costs from wholesale. This can also happen in large residential buildings. These buildings are usually high-end, and they use the best construction materials to make them waterproof and anti-theft. Although the contractor can still wholesale these high-quality materials, the unit cost reduction through wholesale is not enough to make the unit price of materials at the end of construction less than the unit price of materials at the beginning of construction. These good materials are too expensive to achieve the effect of wholesale [13]. Eventually, higher unit price of building materials at the end of construction lead to higher sale price.

3.2 Feature Engineering

Data and features determine the upper limit of machine learning, and models and algorithms could only approach this upper limit. The ultimate purpose of feature engineering is to improve the performance of the model. [14] Therefore, in this part, the goal is to extract features from the original data to the maximum extent for algorithms and models. From a mathematical point of view, feature engineering is to manually design input variables. And data cleaning is the first step. Retaining all the data instead of dropping or changing some of their values is better because the dataset's size is small, and its range is logical. Secondly, the work also drops some redundant features by analyzing the correlation between them in order to avoid the curse of dimensionality.

Before doing analysis, some clean and preprocessing to the original data should be done first. The first thing is to check whether there is any missing value which is displayed as 'NaN' (not a number) in our data. The function ***isnull().any().sum()***in pandas is used to detect the missing values. This function first checks every single element whether it is null or not, then it returns the sum of the element which is null, which makes it more intuitive to list the number of missing values in each column. And the return value is 0. It follows that there are no missing values in the dataset. Besides, before analysis and modeling the data format should be uniform. In this dataset, all the values of features are numbers so that they can be directly used for analysis and model building.

Then the work uses principal component analysis to determine how many features are needed to retain. PCA is usually used for dimensionality reduction of high-dimensional data. It can project the original high-dimensional data to a low-dimensional space and make its variance as large as possible. If a feature of the data (a column of the matrix) is particularly large, the projection will try to approximate the largest feature and ignore the smaller feature. Since we don't know the importance of each feature before modeling, this may result in a lot of missing information. In order to avoid over-capturing some features with large values, we will standardize each feature to make them all in the same range before PCA.There is a normalization api in the SkLearn library.API is shown in Figure 8 below.

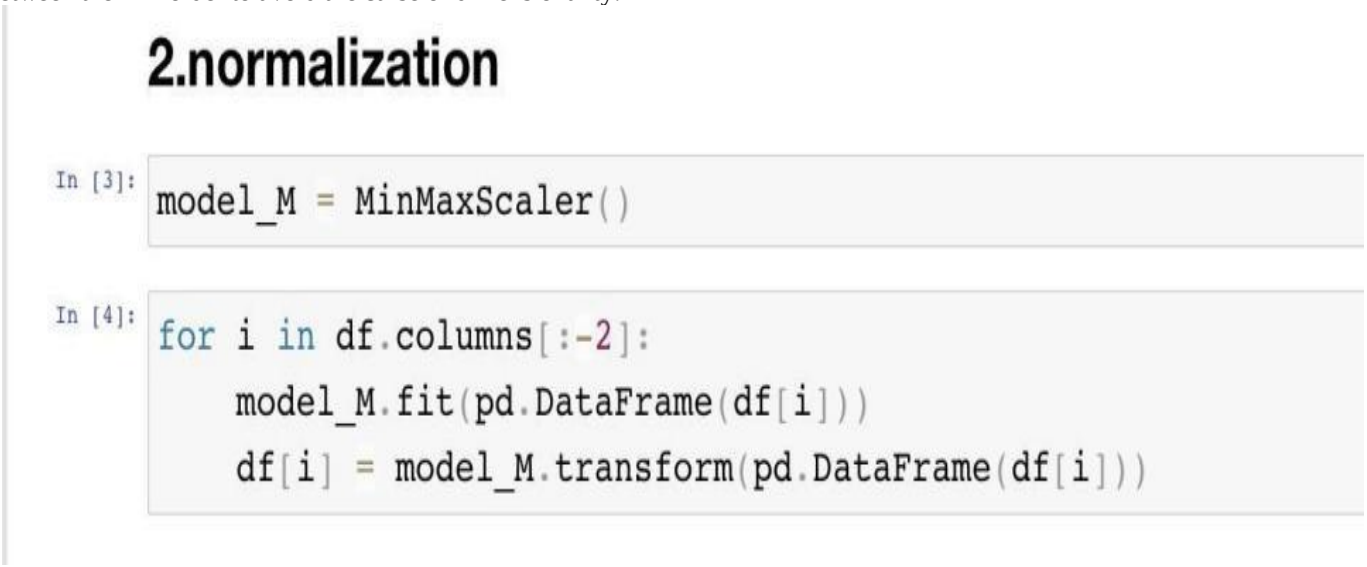


Figure 8. normalization api

Sklearn (sciKit-learn) is a Machine learning tool based on Python. It is built on NumPy, SciPy, Pandas, and Matplotlib.There are six task modules in Sklearn: classification, regression, clustering, dimensionality reduction, model selection and pre-processing, as shown in Figure 9 below. There is also a PCA api in the SkLearn library, if more than 90% of the variance is preserved. So at least eight more features should be preserved based on the results of the code run. The result is shown in Figure 9 below.

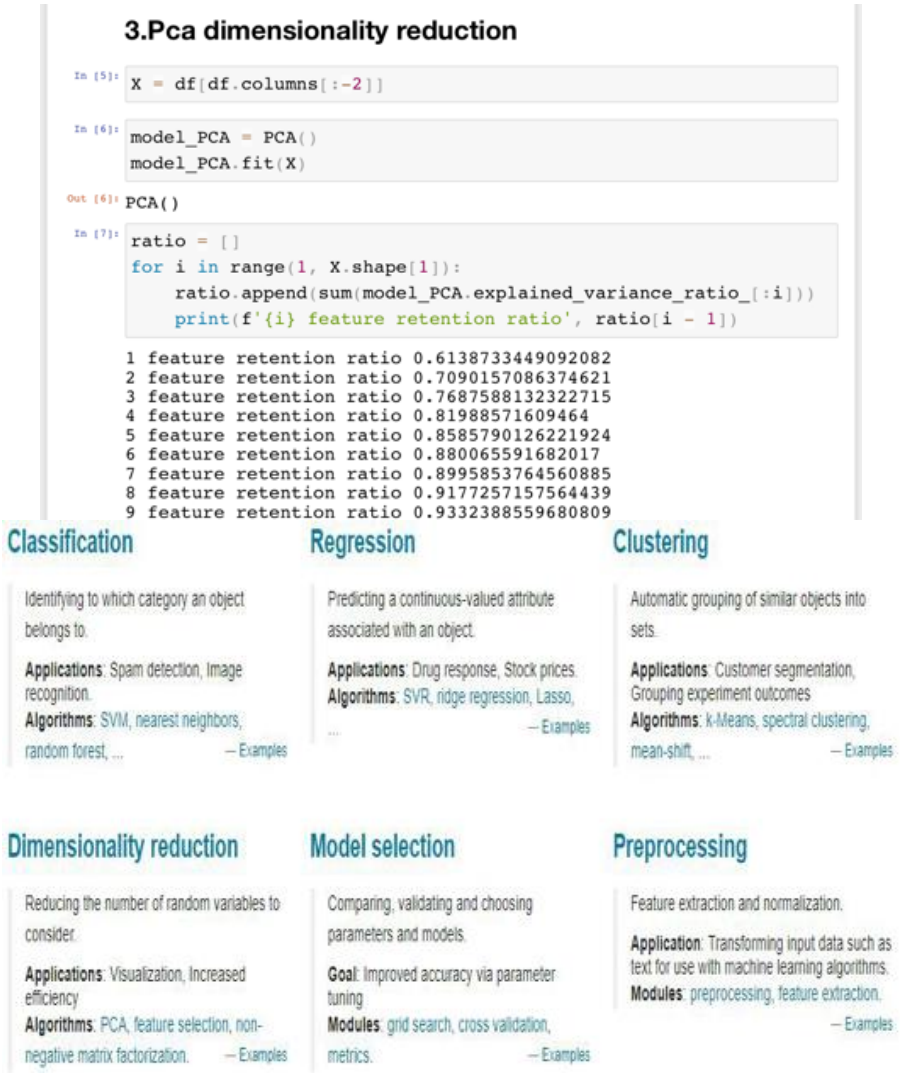


Figure 9. Refer to PCA and the content of SkLearn library

Standardization and normalization are unnecessary since the model used for prediction is a decision tree. However, it should be noted that the PCA dimension reduction used before applying the decision tree algorithm needs to be normalized. In order to find the best, split it searches on each dimension (attribute) a split point which is basically an if clause which groups target values corresponding to instances which have test value less than split value, and on the right the values greater than equal. [15] The criteria are less than or greater than, which means that the real information from the attributes in order to find the split (and the whole tree) is only the order of the values. However, standardization and normalization wouldn't change the original order of our data, which means we will get the same tree. Therefore, standardization and normalization are not necessary.

Due to the curse of dimensionality, the performance of the classifier decreases if there are too many features. In order to drop redundant features to reduce the dimensionality, the work also calculates the correlation index with function. corr**()**(which returns the Pearson correlation coefficient between two features) in pandas between these features to determine how many of them are actually helpful. And we believe that highly correlated features with output results are valuable. High correlation means that this feature can better predict our output. At the same time, the work also needs to avoid multicollinearity. Multicollinearity refers to the distortion or difficulty in estimating the model due to the existence of accurate correlation or high correlation between interpretative variables in linear regression models. If several variables have high correlation to one another, only the most essential one is worth retaining. The features are divided in two parts in the dataset. The work first analyzes the correlation of physical and financial variables in Figure 5. The result is that v3(Lot area) is like v2(total floor area) because they have a high correlation (0.95) and they both reflect the house area, thus v3 is removed. Similarly, feature v4(Total preliminary estimated construction cost based on the prices at the beginning of the project) and v6(Equivalent preliminary estimated construction cost based on the prices at the beginning of the project in a selected base year) contain similar information as v5(Preliminary estimated construction cost based on the prices at the beginning of the project). And

based on Figure 5, v5 had a high correlation (0.78 0.96) with the outputs, so in order to avoid overfitting, v5 is dropped and v4 and v6 are retained. The correlation of the variables are shown in Figure 10 below.

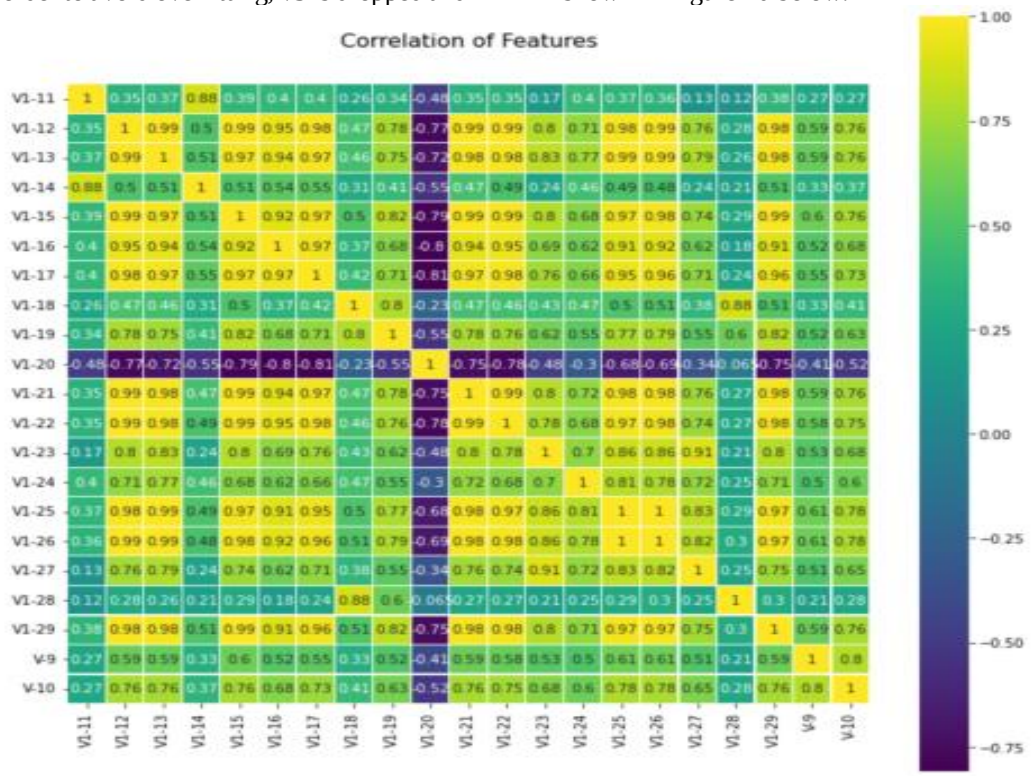


Figure 10. Correlation of features (economic variables and indices)

Examining Figure 10, 9 features have a correlation index which is higher than 0.9 with v29(gold price per ounce), they probably reflect the same economic effects, such as inflation and the fluctuation in supply and demand. Only a few of them is used in the model. Others, which have a relatively low correlation with other features like v14(Total floor areas of building permits issued by the city/municipality), may be related to other influencing factors like the change of policies, thus they are retained for now.

After PCA and preliminary analysis, the work has identified several valuable as well as redundant features. v3(Lot area), v5(Preliminary estimated construction cost based on the prices at the beginning of the project) and v28(Population of the city) should be eliminated first. For the physical and financial variables, v1(Project locality defined in terms of zip codes), v2(Total floor area of the building), v4(Total preliminary estimated construction cost based on the prices at the beginning of the project), v6(Equivalent preliminary estimated construction cost based on the prices at the beginning of the project in a selected base year) and v7(Duration of construction) are closely related to the actual sales prices and actual construction costs. For the economic variables and indices, the work identifies that v12(Building services index (BSI) b for a preselected base year), v13(Wholesale price index (WPI) of building materials for the base year), v15(Cumulative liquidity), v16(Private sector investment in new buildings), v17(Land price index for the base year), v21(The average construction cost of buildings by private sector at the time of completion of construction), v22(The average of construction cost of buildings by private sector at the beginning of the construction), v25(Consumer price index (CPI) in the base year), v26(CPI of housing, water, fuel & power in the base year) should be considered together and only two or three of them would be selected. Other economic variables like v14(Total floor areas of building permits issued by the city/municipality) which may contain other important factors are also retained and should be researched separately.

4. Algorithm and Modeling

First, for the modeling part, a decision tree model was chosen to obtain all possible choices and to accurately figure out which is the best option. Using the scikit-learn library based on Python, which is a machine learning library with implications of classification, regression and clustering algorithms. In order to get an accurate model, regression trees were used to make the predictions since classification methods would result in low accuracy scores. A regression algorithm is more suitable for estimating the construction and sales price of a building within a certain range. Using regression means r2 is used to determine the model score. r2 is determined by the following equation where RSS is the first sum of errors and TSS is second sum of errors. For r2, the highest possible score is 1, and in a very bad model, the score can also go below 0.

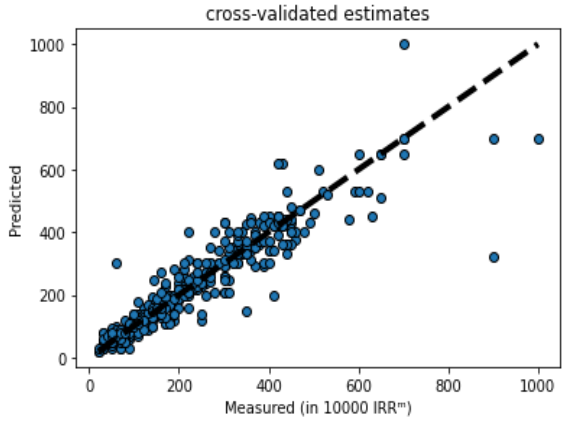


Figure 11. Outputs predicted and actual.

The dataset has 371 rows. They were split into two sets train_test_split from sklearn.model_selection, half for training and half for testing. According to previous research, not only the physical and financial variables can influence the prediction result, but some economic variables like interest rate, loan and cumulative liquidity have a great impact on the construction cost and sale price of the building as well. Therefore, it was decided to use both sets of variables. And only economic variables from time period 1 were added since the economic variables' distributions from 5-time lags are like each other by visualizing them. By doing this, our r2 score using the r2_scorer from sklearn metrics was able to achieve higher than 0.9 out of 1. In Figure 11, the predicted result was close to actual when using both 2 sets of variables. According to the feature engineering, 8 important features were selected to build the model. They are: the total floor area(v2), the total preliminary estimated construction costs based on prices at the beginning of the project (v4), equivalent preliminary estimated construction cost based on the prices at the beginning of the project in a selected base year (v6), the duration of construction (v7), the total floor areas of building permits issued by the city/municipality(v14), Cumulative liquidity (v15), the average construction cost of buildings by private sector at the time of completion of construction(v21), gold price per ounce (v29). Using these variables and some looping, the highest tree depth is 10. Using more variables increases the amount of time the model must branch before giving out a result. This is likely due to the increase in variables, causing more splits to occur before getting the result. Figure 12 below shows the predicted results of the model.

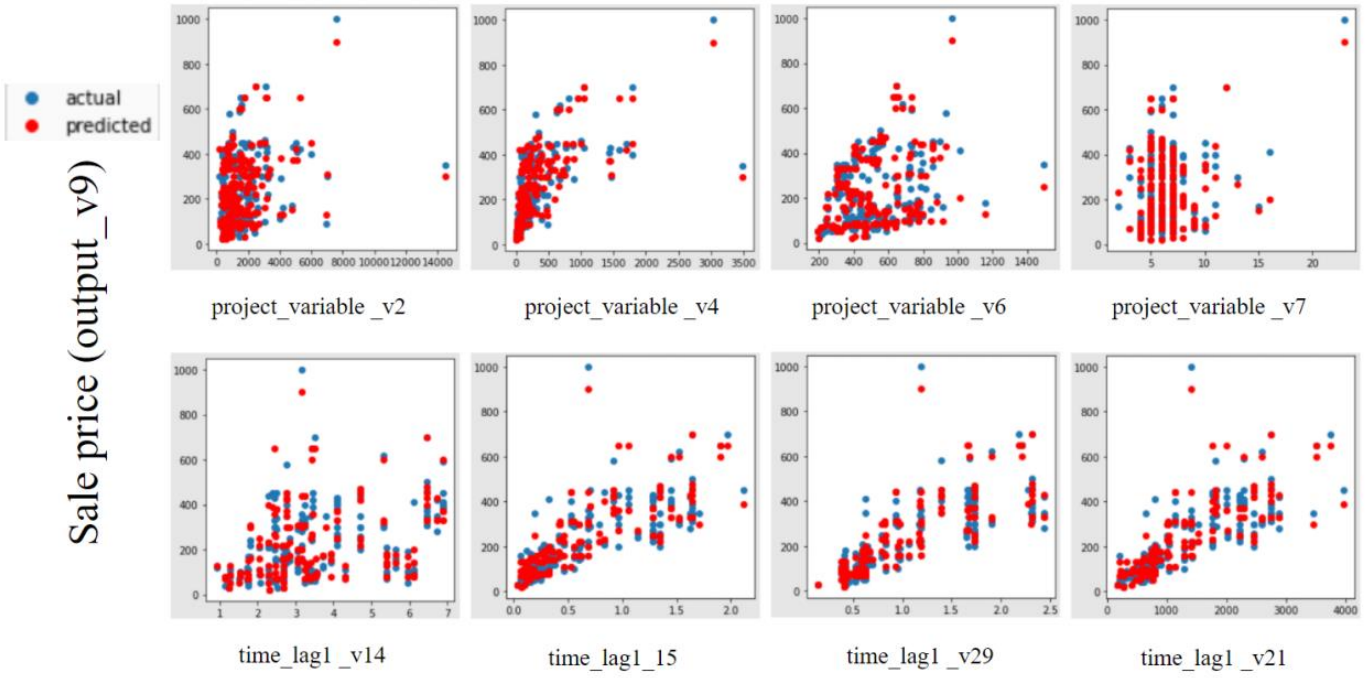


Figure 12. Predicted and actual results of models

5. Conclusion

After analyzing the input variables' influences on the two outputs by visualization, during which some plotting functions in python was used to make pictures embodying the variables relationships, feature engineering work was done to calculate the correlation between each variable and the outputs in order to find out which variable is suitable for building the model. Then a decision tree model was constructed based on the CART algorithm, and the model was able to get a high degree of accuracy when predicting the construction cost and sale price of the residential building in order to help decide whether the residential building is worthy to be built.

Although now the model's accuracy can get above 0.90, the world is always changing, so in order to improve the model, a better method should be used to build the model. The

AdaBoostRegressor is a good tool to build prediction models. An AdaBoostRegressor is a meta-estimator that begins by fitting a regressor on the original dataset and then fits additional copies of the regressor on the same dataset but where the weights of instances are adjusted according to the error of the current prediction. As such, subsequent regressors focus more on difficult cases.
AdaBoostRegressor has five basic parameters:
(1)base_estimator: The base estimator from which the boosted ensemble is built.
(2)n_estimators: The maximum number of estimators at which boosting is terminated. In case of perfect fit, the learning procedure is stopped early.
(3)learning_rate: Weight applied to each regressor at each boosting iteration. A higher learning rate increases the contribution of each regressor. There is a trade-off between the learning_rate and n_estimators parameters.
(4)loss{'linear', 'square', 'exponential'}: The loss function to use when updating the weights after each boosting iteration.
(5)random_state: Controls the random seed given at each base_estimator at each boosting iteration. Thus, it is only used when base_estimator exposes a random_state. In addition, it controls the bootstrap of the weights used to train the base_estimator at each boosting iteration. Pass an int for reproducible output across multiple function calls. See Glossary.[16]

Compared to the CART algorithm used now, AdaBoostRegressor often has a higher accuracy because it combines a lot of weak classifiers(like the CART algorithm) and becomes a strong classifier by iteration. Qiu Zhi Niao(2019) published an article using AdaBoostRegressor and Decision Tree to predict the sale prices of the buildings and find out that after dozens of iterations, the AdaBoostRegressor's accuracy will be a lot lower than the Decision Tree model, which proves that AdaBoostRegressor can help us build a better model. [17]

References

[1] Chen, B. (2021) Analysis on Influencing Factors of house price in medium-sized cities. Northern economy and trade.
[2]Zhang, M and Liu, Y. (2021) Analysis on the driving factors of urban real estate price trend and fluctuation in China -- Empirical Evidence from 31 provinces and 70 large and medium-sized cities in China. Nanjing Social Sciences..
[3]Pan, T. (2021) Review on Influencing Factors of House price. Guangxi quality supervision guide.
[4]Waiwai, 2017. House price prediction. https://zhuanlan.zhihu.com/p/29104647.
[5]CSZhengyang, 2018. [project practice] house price prediction model based on random forest algorithm. https://blog.csdn.net/sunyaowu315/article/details/82982989.
[6]Kaggle administrator, 2016. House Prices - Advanced Regression Techniques. https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data.
[7]Hodgson, Lance. "Construction Delays Due to Weather: Manage Weather Delays like a Pro." *Sitemate*, 21 Mar. 2019, https://sitemate.com/us/resources/articles/commercial/construction-delays-due-to-weather/.
[8]Hinesman, Genae Valecia. "10 Reasons Why Projects Go above Budget." *Small Business - Chron.com*, Chron.com, 26 Oct. 2016, https://smallbusiness.chron.com/10-reasons-projects-above-budget-40886.html.
[9]"What Are the Different Types of Estimates?" *Agiled.app*, 13 Sept. 2021, https://agiled.app/hub/estimates/different-types-of-estimates/.
[10]"Quantity Estimation and Scheduling in Construction Projects." *Techture*, 6 Apr. 2021, https://www.techture.global/quantity-estimation/.
[11] *Hendrickson, C., Au, T., Hendrickson, C., & Au, T. (2008). Cost Estimation. In Project Management for construction: Fundamental Concepts for owners, engineers, architects, and builders. essay, Carnegie Mellon University.*
[12]Afary , Janet. "Iranian Revolution." *Encyclopædia Britannica*, Encyclopædia Britannica, Inc., https://www.britannica.com/event/Iranian-Revolution.
[13]Pitt, Jeson. "6 Ways Contractors Can Save Money on Buying Building Materials." *6 Ways Contractors Can Save Money on Buying Building Materials*, https://www.cross-check.com/blog/6-ways-contractors-can-save-money-on-buying-building-materials.
[14]Pranjal Pandey."Data preprocessing: Concepts"*towards data science*,25 Nov.2019 https://towardsdatascience.com/data-preprocessing-concepts-fa946d11c825.
[15]scikit-learn developers. "6.3 Preprocessing data, 6.3.3Normalization"*scikit-learn*, https://scikit-learn.org/stable/modules/preprocessing.html#preprocessing-data.
[16]scikit-learn official. sklearn.ensemble.AdaBoostRegressor. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostRegressor.html
[17]Qiuzhiniao,2019. AdaBoost algorithm: actual combat. https://zhuanlan.zhihu.com/p/68770891.